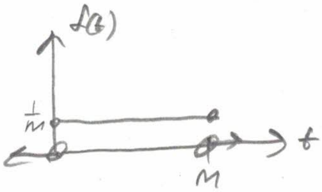
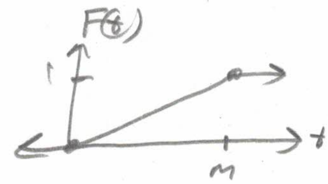
Lec 17 MATH 343

Consider T to be a continuous survival time i.e. $S_T \subseteq (0,\infty)$

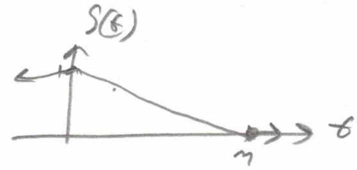
Back to MATH 340... let $T \sim U(0,m) := \frac{1}{m}\mathbb{1}_{t\in[0,m]}$



$$F(t) = \begin{cases} 0 & \text{if } t \le 0 \\ \frac{t}{m} & \text{if } t \in [0,m] \\ 1 & \text{if } t > m \end{cases}$$

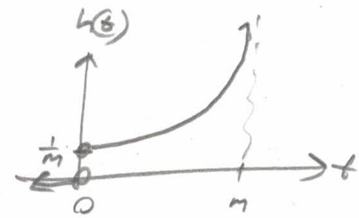


$$S(t) = \begin{cases} 1 & \text{if } t < 0 \\ 1-\frac{t}{m} & \text{if } t \in [0,m) \\ 0 & \text{if } t > m \end{cases}$$



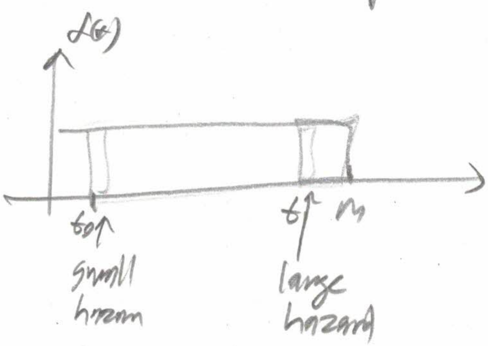
Define the ~~hazard rate~~ hazard function/ $h(t)$ of the rv as

$$h(t) := \underbrace{\frac{f(t)}{S(t)}}_{\substack{>0 \\ \text{on } S_T}} = \frac{\frac{1}{m}\mathbb{1}_{t\in[0,m]}}{\frac{m-t}{m}\mathbb{1}_{t\in[0,m]}} = \begin{cases} \frac{1}{m-t} & \text{if } t \in [0,m] \\ 0 & \text{if } t < 0 \\ \text{undefined} & \text{if } t > m \end{cases}$$



You can think about this as how likely the rv is to realize in the next short period given that it made it this far. It is unique and defines the rv like $f(t), F(t), S(t)$.

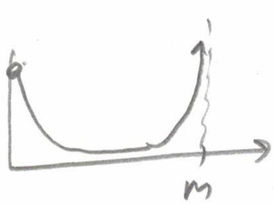
You can see it from the density alone:



$T \sim \text{Exp}(\lambda) := \lambda e^{-\lambda t}\mathbb{1}_{t>0}$, $S(t) = e^{-\lambda t}$

$\Rightarrow h(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$, constant hazard due to memorylessness property.

Consider a rv T with $S_T = [0,m]$ and a bathtub hazard:



This is human lifespan, lightbulb failure etc. Since there is a short period at the beginning which determines long-term viability. Then as it ages, there's a maximum expiration date.

Some identities involving hazard rates:

$$h(t) = -\frac{d}{dt}\left[\ln(S(t))\right] \overset{\text{proof:}}{=} -\frac{\frac{d}{dt}[S(t)]}{S(t)} = -\frac{\frac{d}{dt}[1-F(t)]}{S(t)} = \frac{f(t)}{S(t)}$$ ✓

$\Downarrow$

$$\frac{d}{dt}\left[\ln(S(t))\right] = -h(t)$$ looks like $g(u) = \frac{d}{du}[G(u)]$

$\Downarrow$ second F.T.C. (Newton-Leibniz thm.)

$$\ln(S(b)) - \ln(S(a)) = \int_a^b -h(u)\,du \xrightarrow{\text{let } a=0,\ b=t} \ln(S(t)) - \ln(S(0)) = -\int_0^t h(u)\,du$$

$S(0) = 1$ for all survival distr

$$\Longrightarrow \ln(S(t)) = -\int_0^t h(u)\,du \Rightarrow S(t) = e^{-\int_0^t h(u)\,du}$$ (survival-hazard identity)

Assume $h(t) = h_0(t)\, e^{\beta_1 x_1 + \dots + \beta_p x_p} = h_0(t)\, \underbrace{e^{\vec{x}\vec{\beta}}}_{\text{not a function of } t}$

Note: there is no intercept here as the intercept term "e^{β_0}" is already subsumed inside $h_0(t)$.

This is called a prop. hazard model. why?

Consider $x_{1} = x_{1a} \Rightarrow h_a(t) = h_0(t)\, e^{\beta_1 x_{1a} + \beta_2 x_2 + \dots + \beta_p x_p}$

Consider $x_1 = x_{1b} \Rightarrow h_b(t) = h_0(t) = e^{\beta_1 x_{1b} + \beta_2 x_2 + \dots + \beta_p x_p}$

$$\frac{h_a(t)}{h_b(t)} = \frac{\cancel{h_0(t)}}{\cancel{h_0(t)}} \frac{e^{\beta_1 x_{1a}}}{e^{\beta_1 x_{1b}}} \cdot \cancel{\frac{e^{\beta_2 x_2 + \dots + \beta_p x_p}}{e^{\beta_2 x_2 + \dots + \beta_p x_p}}} = e^{\beta_1 (x_{1a} - x_{1b})} = \text{constant wrt to time } t.$$

The feature values results in multiplicative constants on the hazard rate for all t. If $x_{1a} - x_{1b} = 1$, e^{β_1} is the constant multiplier

If we specify $h_0(t) \Rightarrow$ we know $f(t)$ and we can use maximum likelihood. This not interesting since we could've just worked with the likelihood directly e.g. assume

$Y_i \overset{ind}{\sim} \text{Weibull}(e^{\vec{x}_i \vec{\beta}})$.

David Cox

The cool thing about this model is figured out a way to allow $h_0(t)$ to be arbitrary, hence you are not making an explicit parametric assumption; you're making a "semi-parametric" assumption.

By the survival-hazard identity,

$$S(t) = e^{-\int_0^t h_0(u)\, e^{\vec{x}\vec{\beta}}\, du}$$

By the definition of hazard function,

$$h(t) := \frac{f(t)}{S(t)} \Rightarrow f(t) = h(t)\,S(t) = h_0(t)\, e^{\vec{x}\vec{\beta}}\, e^{-\int_0^t h_0(u)\,du\; e^{\vec{x}\vec{\beta}}}$$

We can now write the likelihood. Assume WLOG data is ordered $y_1 < y_2 < \dots < y_n$ the times and the matrix $X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_n \end{bmatrix}$ in that order as well

$$\mathcal{L}(\vec{\beta}, \underset{\substack{\text{nuisance}\\ \text{parameter:}\\ \text{an entire}\\ \text{function}}}{h_0} ; X, \vec{y}) = \prod_{i=1}^{n} \underbrace{h_0(y_i)}_{h_i}\, e^{\vec{x}_i \vec{\beta}}\, e^{-\overbrace{\int_0^{y_i} h_0(u)\,du}^{H_i}\; e^{\vec{x}\vec{\beta}}}$$

$$\Rightarrow \mathcal{L}(\vec{\beta}, \underbrace{h_1, \dots, h_n, H_1, \dots, H_n}_{\substack{\text{all nuisance parameters}\\ \text{since } h_0(t) \text{ was}\\ \text{unknown}}} ; X, \vec{y}) = \prod_{i=1}^{n} h_i\, e^{\vec{x}_i \vec{\beta}}\, e^{-H_i e^{\vec{x}_i \vec{\beta}}}$$

Make assumptions: (1) All y_i's uniquely-valued (2) $H_i = \int_0^{y_i} h_0(u)\,du \approx h_0(y_1) + h_0(y_2) + \dots + h_0(y_i)$

$$= h_1 + h_2 + \dots + h_i$$

$$= \sum_{l=1}^{i} h_l$$

Kind of like a Riemann Integral Approximation